Some tossing experiments with biased coins

Moors, J.J.A.

Publication date:
1985

REEKS TER DISCUSSIE

No. 85.11

Some tossing experiments with biased coins

J.J.A. Moors

FACULTEIT DER ECONOMISCHE WETENSCHAPPEN

SUBFACULTEIT DER ECONOMETRIE

Some tossing experiments with biased coins

J.J.A. Moors

Summary Several 'coins' were purposively prepared to have biases as large as possible. Nevertheless, no significant biases were found in fairly extensive tossing experiments. An application to Bayesian analysis is given.

1. Introduction

Coin tossing undoubtedly is one of the most familiar statistical experiments. A single toss can be described formally as a Bernoulli experiment with the two possible outcomes heads and tails. For a given coin the corresponding probability distribution is determined by p: = P(heads). In general, p is an unknown parameter that has to be estimated from a fixed number of consecutive tosses.

The classical estimator for p is the observed relative frequency of heads. However, this estimation problem is particularly suited for a Bayesian approach, because a lot of information is available in advance. At a theoretical level, this information concerns symmetry: during the production much care is taken to guarantee that coins show certain symmetries. Further, much empirical evidence is at hand, since coins are being tossed very frequently. A Bayesian statistician considers the specific value of p as the outcome of a random experiment (random selection from the set of Dutch coins) and uses his prior information to specify a probability distribution for the underlying random variable $\underline{p}$. (Note that random variables will be underlined here.)

To describe properly the tails of the prior distribution, it is of importance to know what biases coins may have. The theoretical maximum of $|p-\frac{1}{2}|$ of course equals $\frac{1}{2}$ and is attained for coins with two identical surfaces. Since such false coins are easily recognized, they will be excluded from now on. On the other hand, to allow larger biases, the definition of a coin will not be made too restrictive: in the sequel all objects with the measures of a coin and two distinguishable surfaces will be called coins.

The question which of the just defined coins has maximum bias is essentially a physical one. Since bias is caused by shifting the centre of gravity along the axis of the cylinder, in my opinion maximum bias will be achieved in a circle-symmetric coin. The following three models are serious candidates for having maximum bias. All three have the measures of a 'rijksdaalder', the largest regular Dutch coin; their diameters equal 29 mm, their heights 1.8 mm.

Model A is a normal rijksdaalder with a circular hole drilled into the heads side halfway through; see Figure 1. The diameter of the hole is 20 mm, its depth 0.9 mm.

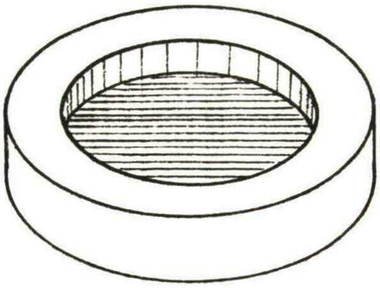

Figure 1

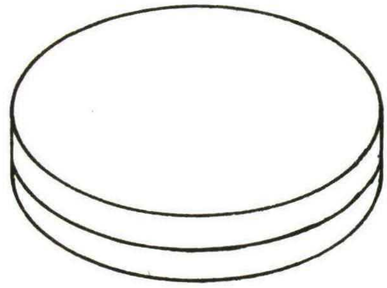

Figure 2

Models B and C are not real coins. Model B consists of two equally thick layers, one of plastic and one of lead; see Figure 2. The plastic side will be called heads.

Model C is similar to model B, except that aluminium is used in stead of plastic; the aluminium side will be called heads.

Note that all three models can be expected to have a positively biased p, since the less heavier side will come on top more often. Mr. Nico L. Willemse of the Psychological Laboratory of Tilburg University was kind enough to prepare two coins of each of the three models.

All tossing experiments took place under very similar conditions; the precise instructions handed to the experimentators one shown in Appendix 1. The results were noted on standard formulars; Appendix 2 offers an example. As coins were flipped into the air from a horizontal starting position, it might matter which side was on top at the start. Therefore, this starting position was changed after one halve of the number of tosses with each coin. This number equalled 6 000 for model A and 3 000 for models B and C each. I am very grateful to my childeren, Sinbad and Tamar Moors, for performing a large part of the experiments. I hate to add that their efforts where not purely for the sake of science: their rate was 1 cent per throw, later raised to 1.5 cent.

The main results are summarized in Table 1; the main findings for model A were reported earlier in the accompanying Stelling 2 to MOORS 1985.

Table 1 Relative frequencies of heads in coin tossing experiments

Model	Starting position		Total
	Heads on top	Tails on top	
A	0.507	0.514	0.510
B	0.497	0.507	0.502
C	0.499	0.495	0.497

Section 2 presents a more detailed account of the results with model A; several tests are applied. Sections 3 and 4 give similar analyses for models B and C, respectively. The findings are discussed in the final section 5; there an application to Bayesian statistics is given as well.

2. Empirical results for model A

Table 2 presents the observed frequencies of heads among the 6,000 tosses with the excavated rijksdaalder. Figure 3 shows the total series of heads and tails in more detail.

Table 2 Observed frequences of heads; model A

	Starting position		Total
	Heads on top	Tails on top	
no. of tosses	3 000	3 000	6 000
freq. of heads	1 520	1 542	3 062
relative freq.	0.507	0.514	0.510

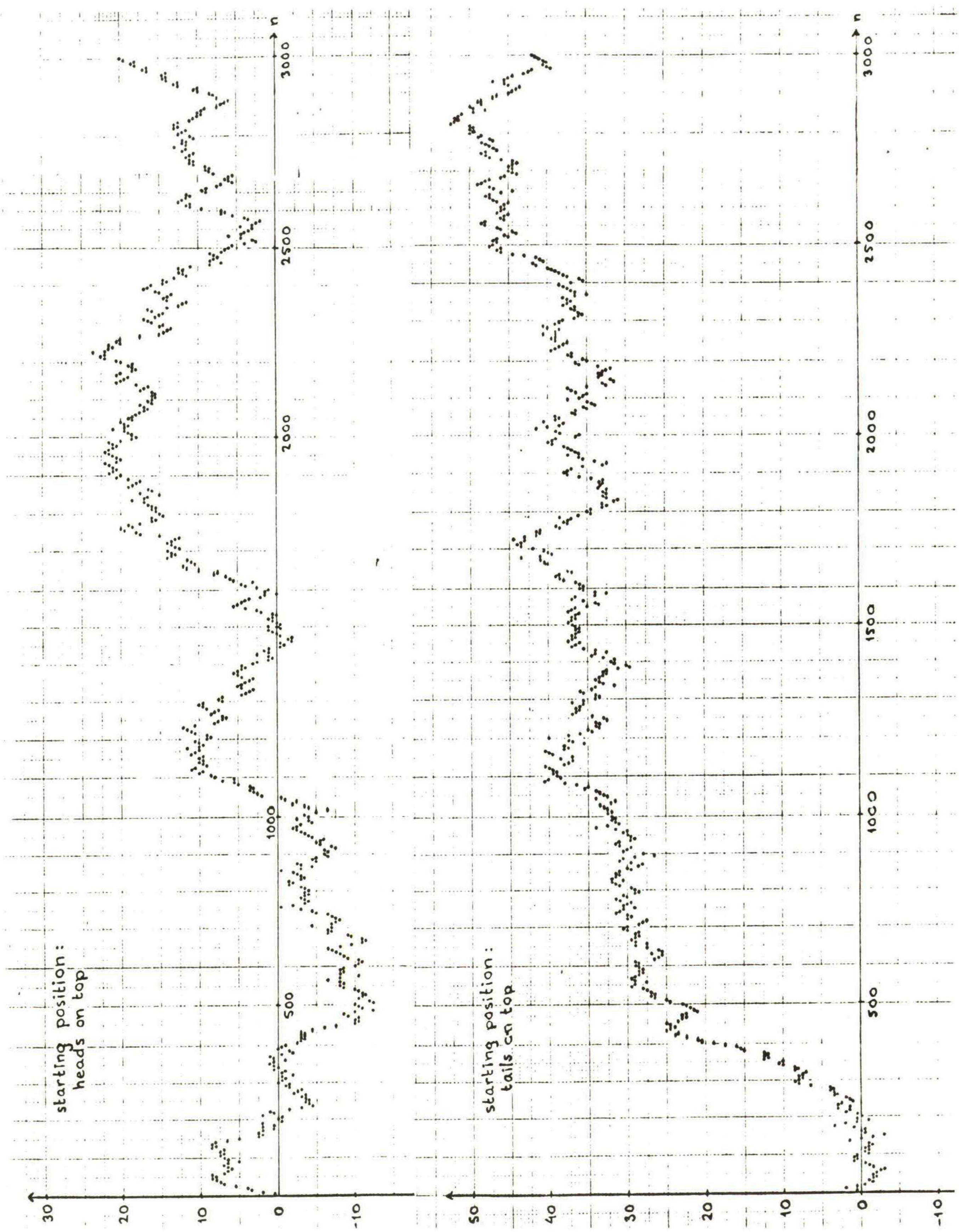

Figure 3 Number of tosses (n) and number of outcomes heads exceeding n/2; excavated rijksdaalder

The results imply that for both starting positions the null-hypothesis $p = \frac{1}{2}$ can stand a one-sided test at 5% level. The findings for the two starting positions show no significant difference and hence may be pooled. Even the pooled frequencies are not significantly different from $\frac{1}{2}$ at 5% level. So, only a slight bias results (if any), although the coin has been tampered with very thoroughly.

Next, as a check on randomness the frequencies of heads were noted for any 25 consecutive tosses. Table 3 shows the results; n denotes the number of heads in 25 tosses; for comparison the theoretical frequencies based on the binomial distribution $B(25,\frac{1}{2})$ are added.

Table 3 Frequencies of heads per 25 tosses; model A

n	Starting position		Theoretical freq. ($p=\frac{1}{2}$)
	Heads on top	Tails on top	
≤ 8	5	3	6.5
9	8	9	7.3
10	12	9	11.7
11	13	11	15.9
12	21	20	18.6
13	12	20	18.6
14	22	17	15.9
15	12	14	11.7
16	6	11	7.3
≥ 17	9	6	6.5
Total	120	120	120

The χ^2-values for the goodness-of-fit test are 6.5 (7.1) for the cases heads (tails) on top. Comparison with the critical value $\chi^2_{9;0.95} = 16.9$ shows that the data pass this test. (For the pooled data the χ^2-value equals 8.21.)

As a further test on randomness runs were counted; a run of heads is a consecutive series of outcomes heads, preceeded and followed by an outcome tails. Table 4 shows the frequencies of runs of different lengths. The last column gives the expected frequencies for fair coins; see for example FISZ 1963, Ch. 11.

Table 4 Frequencies of runs; model A

Runlenght	Starting position: Heads on top, Heads	Heads on top, Tails	Tails on top, Heads	Tails on top, Tails	Theoretical freq. ($p=\frac{1}{2}$)
1	408	404	356	390	375.2
2	175	199	204	180	187.6
3	92	81	97	90	93.8
4	44	48	38	48	46.9
5	26	18	26	23	23.5
6	13	7	10	8	11.7
7	8	9	13	2	5.9
$\geq$ 8	5	5	6	8	5.9
Subtotal	771	771	750	749	750.5
Total	1542		1499		1501

The maximum runlength observed was 13; it occurred thrice.

Two different tests were applied to these data. The first concerns the total number of runs; since this number has the normal distribution N(1501,750) approximately, no significant deviation can be detected for both starting positions. Secondly, the distribution of run lengths can be tested by a conditional χ^2-test, the condition being the observed subtotal number of runs. To apply this test the theoretical frequency distribution has to be adapted proportionally to make the subtotal number of runs agree with the observed number. The calculated values of the test statistic were 4.5, 8.8, 13.3 and 5.5 for the four successive cases of Table 4. Comparison with $\chi^2_{7;0.95} = 14.1$ shows that all observed distributions of run lengths agree with the theoretical one.

3. Empirical results for model B

For the coin consisting of plastic (heads) and lead, Table 5 prsents the observed frequencies of heads.

Table 5 Observed frequencies of heads; model B

	Starting position: Heads on top	Tails on top	Total
no. of tosses	1 500	1 500	3 000
freq. of heads	745	760	1 505
relative freq.	0.497	0.507	0.502

Exactly the same conclusion can be drawn as from Table 2.

Table 6 surveys the number n of heads in any 25 consecutive tosses with this model.

Table 6 Frequencies of heads per 25 tosses; model B

n	Starting position: Heads on top	Tails on top	theoretical freq. ($p=\frac{1}{2}$)
$\leq$ 9	11	9	6.9
10	5	2	5.9
11	5	7	7.9
12	9	7	9.3
13	8	11	9.3
14	9	12	7.9
15	6	7	5.9
$\geq$ 16	7	5	6.9
Total	60	60	60

Since the χ^2-values are 4.0 for heads on top and 7.1 for tails on top, the data pass the goodness-of-fit test ($\chi^2_{7;0.95} = 14.1$). For the pooled data the χ^2-value equals 8.0.

The analogon to Table 4 is presented below.

Table 7 Frequencies of runs; model B

Runlength	Starting position: Heads on top, Heads	Heads on top, Tails	Tails on top, Heads	Tails on top, Tails	Theoretical freq. ($p=\frac{1}{2}$)
1	192	190	164	182	187.6
2	93	91	96	104	93.8
3	43	46	54	36	46.9
4	24	24	30	15	23.5
5	15	11	15	14	11.7
6	4	9	3	2	5.9
≥ 7	5	5	4	12	5.9
Subtotal	376	376	366	365	375.3
Total	752		731		750

The maximum runlength was 14; it occurred once. The observed totals are not significant. The (conditional) χ^2-values are 2.1, 2.0, 8.8 and 16.5 respectively; since $\chi^2_{6;0.95} = 12.6$ only the runs of tails (tails on top) show a significantly different behaviour.

4. Empirical results for model C

For the coin consisting of aluminium (heads) and lead, Table 8 shows the observed frequencies of heads; Table 9 is the counterpart of Tables 3 and 6.

Table 8 Observed frequencies of heads; model C

	Starting position: Heads on top	Starting position: Tails on top	Total
no. of tosses	1 500	1 500	3 000
freq. of heads	748	743	1 491
relative	0.499	0.495	0.497

Table 9 Frequencies of heads per 25 tosses; model C

n	Starting position: Heads on top	Starting position: Tails on top	Theoretical freq. ($p=\frac{1}{2}$)
$\leq$ 9	5	7	6.9
10	9	8	5.9
11	9	7	7.9
12	9	10	9.3
13	8	9	9.3
14	4	5	7.9
15	10	7	5.9
$\geq$ 16	6	7	6.9
Total	60	60	60

All data pass the appropriate tests; the χ^2-values are 7.4 (2.3) for heads (tails) on top.

Observed runs of consecutive outcomes heads and of consecutive tails are presented in the next table.

Table 10 Frequencies of runs; model C

Runlength	Starting position: Heads on top, Heads	Heads on top, Tails	Tails on top, Heads	Tails on top, Tails	Theoretical freq. ($p=\frac{1}{2}$)
1	189	209	178	187	187.6
2	97	83	98	83	93.8
3	50	36	41	40	46.9
4	22	18	28	33	23.5
5	13	13	10	11	11.7
6	5	11	5	5	5.9
$\geq$ 7	4	9	7	8	5.9
Subtotal	380	379	367	367	375.3
Total	759		734		750

The maximum runlength observed was 10, which occurred once. The observed total numbers of runs are not significant. The (conditional) χ^2-values are 1.3, 13.4, 2.7 and 7.0, so only the runs distribution of tails (heads on top) deviates significantly from the theoretical distribution.

5. Discussion and Bayes'approach

In the foregoing sections several tests were applied to check whether the numbers of heads and tails, observed for three specially prepared coins, followed a binomial deitribution with probability $\frac{1}{2}$. Two of these tests lead to significant results. However, since 24 tests in total were applied (with level 5%), this number of significant results can easily occur even if all null hypotheses were true.

The general conclusion from the extensive experiments therefore is that all three coins still can be regarded as unbiased, despite the fact the coins were purposively designed to have as large a bias as possible! A fortiori, any regular coin will have a very small bias (if any).

This can be expressed more precisely by constructing a prior distribution for the probability p of heads for regular coins. Such prior distribution expresses the statistician's general knowledge about p before performing the experiments. The parameter p is now viewed as a random variable (the probability of heads for a randomly chosen coin). Only normal distributions with mean $\frac{1}{2}$ will be taken into consideration:

(1) $\underline{p} \sim N(\frac{1}{2}, \sigma^2)$

In view of the above results rather conservative values for the standard deviation σ would be 0.02 or 0.05, while 0.1 is definitely too large.

The Bayesian analysis now proceed by updating the prior distribution by means of the data. Assume that n ($\geq$ 25) tosses are preformed with a given coin and denote the (random) fraction of heads by $\underline{f}$. Than for fixed p the distribution of $\underline{f}$ is approximately given by

(2) $\underline{f} \sim N(p, (4n)^{-1})$

where p(1-p) has been approximated by $\frac{1}{4}$. From (1) and (2) the conditional distribution of $\underline{p}$, given the outcome f of $\underline{f}$, can be derived (see for example DEGROOT 1970). It is called the posterior distribution of $\underline{p}$ and reflects both the statitician's prior knowledge and the information provided by the data. This (conditional) posterior distribution is given by

(3) $\underline{p} \mid f \sim N(\mu_1, \sigma_1^2)$

The posterior mean μ_1 is a weighted mean of the observed fraction f and the prior mean $\frac{1}{2}$ with weights reciprocal to the respective variances:

(4) $\mu_1 = w/2 + (1-w)f$

(5) $w := \dfrac{1/(4n)}{\sigma^2 + 1/(4n)} = \dfrac{1}{4n\,\sigma^2 + 1}$

Further, the posterior variance σ_1^2 is the harmonic mean of the variances σ^2 and 1/(4n):

(6) $\sigma_1^2 = \dfrac{\sigma^2}{4n\,\sigma^2 + 1}$

These results have an obvious intuitive interpretation.

A central quantity appears to be $4n\,\sigma^2$, the ratio between the variances of p and f; Table 11 gives some numerical values. Note that w equals the weight of the prior mean by definition, but also the variance reduction factor σ_1^2/σ^2 - cf. (5) and (6).

Table 11 Variance ratio and weight w

$4n\,\sigma^2$	$w = \sigma_1^2/\sigma^2$	$4n\,\sigma^2$	$w = \sigma_1^2/\sigma^2$
1/99	0.99	99	0.01
1/19	0.95	19	0.05
1/9	0.9	9	0.1
1/3	0.75	3	0.25
1	0.5	1	0.5

Prior information and empirical data are equally important if $4n\,\sigma^2 = 1$; hence, the prior information is 'worth' 100 tosses for $\sigma = 0.05$ and even 625 tosses for $\sigma = 0.02$.

As final numerical example Figure 4 shows the prior distribution with $\sigma = 0.02$ as well as the posterior distributions for $f = 0.52$, calculated from 625 and 1875 tosses respectively. For the probability of a negatively biased coin the values 50%, 24% and 6.7% can be calculated from these three distributions.

It would be interesting to confront the experimental results reported here with the theoretical value of p that could be derived from a physical model of coin tossing. I hope that some physicist will find this problem challenging enough to construct such a model.

Acknowledgement

I am grateful to B.B. van der Genugten and to J. Kriens for their critical reading of the manuscript.

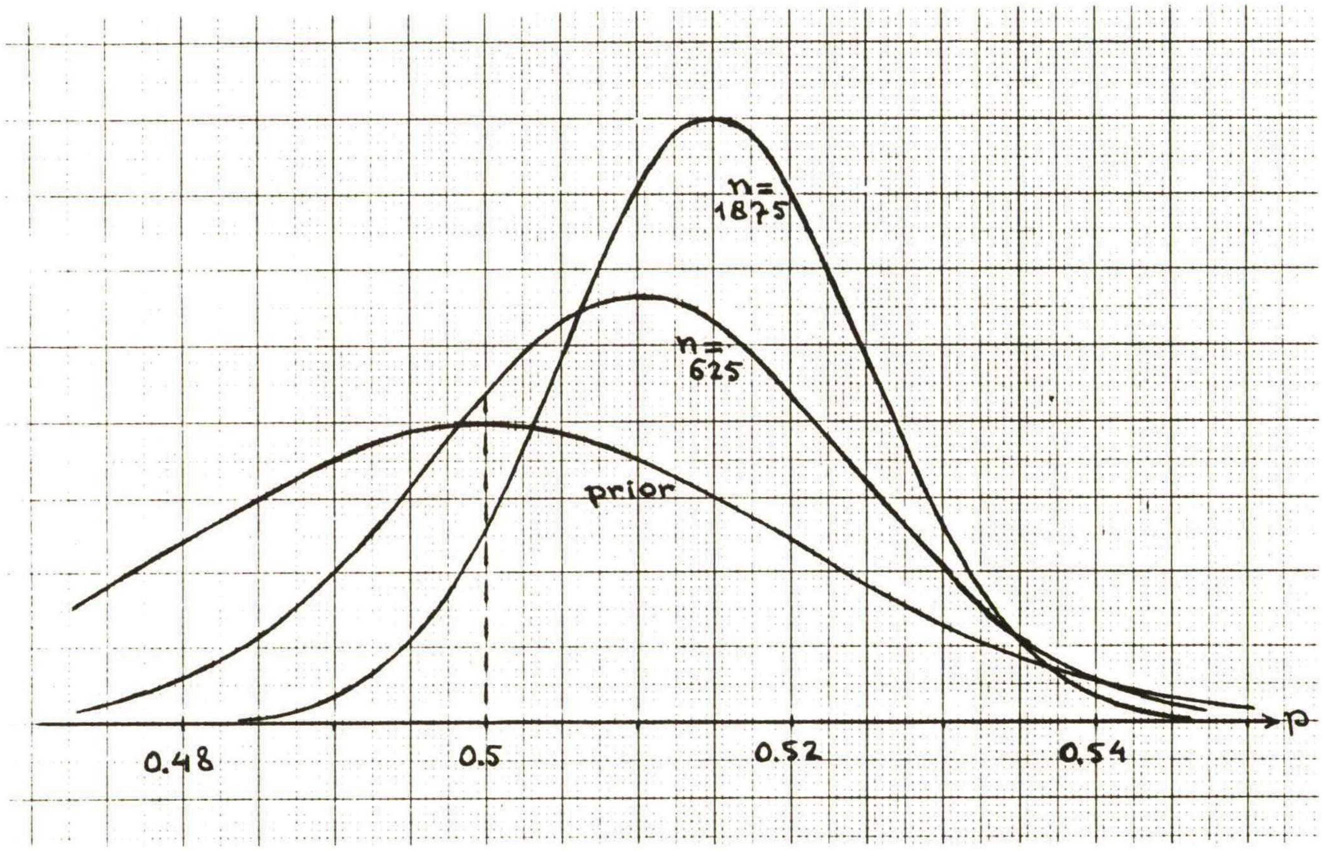

Figure 4 Examples of prior and posterior distributions

References

DEGROOT, M.H. 1970, Optimal statistical decisions, McGraw-Hill, New York.

FISZ, M. 1963, Probability theory and mathematical statistics, Wiley, New York.

MOORS, J.J.A. 1985, Estimation in truncated parameter spaces, Ph.D. thesis, Tilburg University.

Appendix 1 Instructions for tossing excavated rijksdaalder

1. Put the coin horizontally on your thumbnail with the indicated side on top.

2. Flip - standing on the carpet in the living-room - the coin into the air, at least 15 cm. high.

3. Let it fall on the carpet uninterruptedly and look which side is on top.

4. Note the result on the standard formular:

   M : if tails is on top;

   G : if the hole is on top*.

   Mind the right order.

5. Call out 'invalid' whenever an irregularity occurs and disregard the result. Examples of irregularities are:
   - the coin does not ascend high enough or touches the ceiling;
   - when falling down the coin touches some object;
   - the coin rolls off the carpet.

6. A toss is disregarded if one of the persons performing the experiment calls out 'invalid'. Utter this call as soon as possible, in any case before observing the outcome of the toss.

7. Do all this carefully and scrupulously, otherwise the results are of no use, and check each other.

Good luck

* Tails is 'munt' in Dutch and hole is 'gat', hence the codes.

300 worpen met uitgeholde rijksdaalder — 'gat' boven — Blad no. U1

naam: Tamar en Hans — datum: 26/3 — bij elke worp — totaal aantal 'gat' 150

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25		G
G	G	M	M	G	M	G	G	G	G	G	G	G	M	G	G	G	G	M	M	G	G	G	M	G	25	18
G	G	G	G	M	M	G	G	G	M	G	M	G	G	M	M	G	M	G	G	G	M	M	M	G	50	15
M	G	G	M	G	M	M	M	M	G	G	G	M	M	M	M	G	G	M	M	M	G	M	G	G	75	11
G	G	M	M	G	M	M	G	M	G	M	M	G	M	M	M	G	G	G	G	G	G	M	M	G	100	13
G	G	G	M	M	M	M	G	G	M	G	M	G	G	M	M	G	M	G	M	G	M	G	G	G	125	14.
M	M	G	G	M	G	M	G	G	M	M	G	G	M	M	M	M	M	M	G	M	M	M	G	M	150	9
M	M	M	M	M	G	M	M	G	M	M	G	G	G	M	M	G	M	M	G	M	M	M	G	M	175	8
M	G	G	G	G	G	M	M	M	M	G	G	G	G	M	M	M	M	M	M	M	G	M	G	G	200	1~~1~~2
M	M	M	G	G	G	G	G	G	G	M	M	G	M	M	G	M	G	G	M	M	M	M	M	M	225	11
M	G	M	M	M	M	M	M	M	G	G	M	G	M	G	G	M	G	G	M	M	M	G	M	G	250	10
M	M	G	G	G	G	M	G	G	M	M	M	G	M	G	G	G	G	G	M	G	G	M	M	M	275	14
M	M	G	G	M	G	G	M	G	G	M	G	M	G	G	M	G	M	M	G	G	G	M	G	G	300	15
4	7	7	6	6	6	4	7	8	6	6	6	10	5	4	4	8	6	6	5	6	6	3	6	8		150

Appendix 2 Registration formular

IN 1984 REEDS VERSCHENEN

01. P. Kooreman A. Kapteyn	Estimation of Rationed and Unrationed Household Labor Supply Equations Using Flexible Functional Forms	jan.
02. Frans Boekema Leo Verhoef	Lokale initiatieven; Sleutel voor werkgelegenheidsontwikkeling op lokaal en regionaal niveau	febr.
03. J.H.J. Roemen	In- en uitstroom van melkvee in de Nederlandse rundveesektor geschat m.b.v. een "Markov"-model	febr.
04. M.D. Merbis	From structural form to state-space representation	febr.
05. R.H. Veenstra J. Kriens	Steekproefcontrole op ernstige en niet-ernstige fouten (gecorrigeerde versie)	maart
06. Th. Mertens	Kritiek op Habermas' communicatietheorie: een evaluatie van het Gadamer-Habermas-debat en van Habermas' interpretatie van de taalhandelingstheorie. Een onderzoeksverslag	maart
07. P. Bekker A. Kapteyn T. Wansbeek	Measurement error and endogeneity in regression: bounds for ML and IV-estimates	maart
08. B.R. Meijboom	An input-output like corporate model including multiple technologies and "make-or-buy" decisions	april
09. J.J.A. Moors	On the equivalence between cooperative games and superadditive functions	april
10. J. van Mier	Gewone differentievergelijkingen met niet-constante coëfficiënten en partiële differentievergelijkingen (vervolg R.T.D. 83.31)	april
11. W.J. Oomens	Het optimale prijs- en reclamebeleid van een monopolist	april
12. P.A. Verheyen	Een dynamische ondernemingstheorie en de reacties op de overheidspolitiek	mei
13. G.J.C.Th. van Schijndel	Vermogensverschaffersclièntèles in statistische en dynamische ondernemingsmodellen	mei

14. P. Kooreman A. Kapteyn	The effects of economic and demographic variables on the allocation of leisure within the household	mei
15. L. Bosch	Over flexibele produktie-automatisering	juni
16. M. Janssens R. Heuts	On distributions of ratios of dependent random variables	juni
17. J. Plasmans	Specification and estimation of the linkage block of Interplay II (1953-1980)	juni
18. P. Bekker A. Kapteyn T. Wansbeek	Consistent sets of estimates for regressions with correlated or uncorrelated measurement errors in arbitrary subsets of all variables	juni
19. A.L. Hempenius	Dividend policy of large Dutch corporations	juni
20. B.B. van der Genugten K. van der Sloot H.A.J. van Terheijden	Handleiding voor de programma's DATAH en REGAP	juni
21. A.B. Dorsman J. v.d. Hilst	The influence of the calculation-interval on the distribution of returns at the Amsterdam Stock Exchange	juni
22. B.R. Meijboom	Joint and Common Cost Allocation in a Multi-Level Organization	juli
23. Ton J.A. Storcken	Arrow's impossibility theorem on restricted domains	juli
24. E.E. Berns	De Terugtrekking Over politiek en ethiek bij Derrida	juli
25. Chr.H. Kraaijmes	De organisatorische condities voor concrete hulpverlening: een model naar aanleiding van de sociale dienst	juli
26. A.L. Hempenius	The Interpretation of Cross-Sectional Regressions with Variable Constant Terms	aug.
27. J. Kriens J.Th. van Lieshout	Enkele eigenschappen van de kritieke-lijn-methode van Markowitz	aug.
28. A.W.A. Boot	Optimum condities voor een discontinu investeringsprobleem	sept.
29. M.F.C.M. Wijn	Wie liquideert de onderneming?	okt.

30. A.L. Hempenius	Modelling dividend behavior	nov.
31. J.H.J. Roemen	Beslissingsregels voor de investerings- en financieringsactiviteiten van een melkveebedrijf	nov.
32. J. van Mier	Gewone differentievergelijkingen met niet-constante coëfficiënten en partiële differentievergelijkingen (vervolg R.T.D. 84.10)	dec.

IN 1985 REEDS VERSCHENEN

01.	H. Roes	Betalingsproblemen van niet olie-exporterende ontwikkelingslanden en IMF-beleid, 1973-1983	febr.
02.	P. Kort	Aanpassingskosten in een dynamisch model van de onderneming	maart
03.	G.J.C.Th. van Schijndel	Optimale besturing en dynamisch ondernemingsgedrag	maart
04.	J. Kriens J.J.M. Peterse	Toepassing van de regressie-schatter in de accountantscontrole	mei
05.	J. Kriens R.H. Veenstra	Statistical Sampling in Internal Control by Using the A.O.Q.L.-system (revised version of Ter Discussie no. 83.02)	juni
06.	A. van den Elzen D. Talman	A new strategy-adjustment process for computing a Nash equilibrium in a noncooperative more-person game	juli
07.	W. van Eijs W. de Freytas T. Mekel	Automatisering, Arbeidstijd en Werkgelegenheid	juli
08.	A. van Soest P. Kooreman	Nederlanders op vakantie Een micro-economische analyse	sept.
09.	H. Gremmen	Macro-economisch computerspel Beschrijving van een model	okt.
10.	F. van der Ploeg	Inefficiency of credible strategies in oligopolistic resource markets with uncertainty	okt.